

Comparison of Model Evaluation Methods to Develop a Comprehensive Watershed Simulation Model

Misgana Muleta, Ph. D., P.E.

Abstract

Comprehensive environmental models such as the Soil and Water Assessment Tool (SWAT) are becoming an integral part of decision making processes for effective planning and management of natural resources. Before their use as decision making aid, however, models must be properly evaluated to improve their prediction accuracy and reduce the likelihood of making decisions that could lead to undesirable policy outcomes. Model evaluation refers to practices such as quality analysis of input data, sensitivity analysis, calibration and verification, and uncertainty analysis. Many methodologies have been developed for model evaluations over the years. One of the major limitations of the existing model evaluation methods, in particular model calibration methods, is their computational inefficiency, especially when used to calibrate comprehensive watershed simulation models. It may take weeks to months of CPU time, depending on the problem size, to successfully calibrate a comprehensive watershed simulation model on a standard PC. In this study, two sensitivity analysis methods and four calibration methods are used to evaluate SWAT to improve its streamflow prediction accuracy for the Morro Bay watershed located on the central coast of California. Parameter sensitivity analysis was performed using step-wise-regression analysis and the one-factor-at-a time screening method. Calibration was performed using PEST, Genetic Algorithms, the Shuffled Complex Evolution, and the Dynamically Dimensioned Search using observed data from multiple sites in the watershed. The model evaluation methods are compared in terms of their computational efficiency as well as effectiveness to determine “accurate” results. The developed SWAT model can be used to evaluate effectiveness of the Best Management Practices installed in the Morro Bay watershed, and to also prioritize sites where BMPs may be implemented in the future to further improve ecological integrity of the Morro Bay Estuary, which is one of the most important wetlands in California as it supports wide variety of habitats including numerous sensitive and endangered plant and animal species.

Introduction

Watershed simulation models use coupled system of transfer functions that mathematically describe the hydrologic and water quality processes responsible for generating streamflow and contaminants of concern considering unique watershed and stream characteristics. These mathematical functions are defined in terms of measurable input variables as well as parameters that conceptualize some aspects of the watershed processes. The conceptual parameters are often not readily measurable and have to be determined via a series of model evaluation procedure based on observed data. Model evaluation refers to the process of determining model usefulness and estimating the range or likelihood of various outcomes (Matott et al., 2009) and includes procedure such as quality assurance of the input data, sensitivity analysis (SA), model calibration and uncertainty analysis. For model outputs to be useful for applications ranging from academic

research to major policy analysis, models and modeling processes should be scientifically sound, robust, reproducible and defensible (U.S. EPA, 2002). However, since every mathematical model undergoes some level of conceptualization and parameterization, models must pass through rigorous model evaluation procedure before they are utilized as decision making aid in the planning and management of water resources. Many methodologies have been developed for model evaluations over the years (Duan et al., 1992; Beven and Freer, 2001; Muleta and Nicklow, 2005; Tolson and Shoemaker, 2007). The major limitation of the existing model evaluation methods is their computational inefficiency. Depending on the problem size, it could take weeks to months of CPU time to calibrate a comprehensive watershed simulation model thus threatening their practicality for day-to-day applications.

This study explores the application of a comprehensive hydrologic and water quality simulation model known as Soil and Water Assessment Tool (Arnold et al., 1999) to control nonpoint source pollution of a sensitive freshwater estuary on the central coast of California. Sediment is the major contaminant of concern that has been threatening sustainability of the estuary. Two SA and four automatic model calibration methods have been applied to improve accuracy of streamflow predicted for the watershed draining the Bay using SWAT. The SA methods are compared in terms of computational efficiency, consistency of the identified sensitive parameters, and the quality of information available for the modeler to decide what parameters to include in the model calibration stage. The four calibration methods are also compared in terms of their accuracy in reproducing observed streamflow, and in terms of their computational efficiency. This comparative study can assist other modelers select effective and efficient SA and calibration methods among the methods tested in this study. Furthermore, the developed model can be used to identify major sediment sources in the Morro Bay watershed, and to prioritize sites where BMPs may be implemented in the future to improve ecological integrity of the bay. In addition, the model can be used to assess effectiveness of the BMPs that have been implemented in the watershed since mid-1990 to control sedimentation of the bay.

The Watershed Simulation Model

SWAT, a model developed at the USDA's Blacklands research center, is a continuous-time, spatially distributed simulator developed to assist water resource managers in predicting impacts of land management practices on water, sediment and agricultural chemical yields. The model is well suited for large complex watersheds with varying soils, land use and management conditions over long periods of time (Nietsch, et al., 2001; Arnold et al., 1998; ASCE, 1999). SWAT makes use of watershed information such as weather, soil, topography, vegetation, and land management practices to simulate watershed processes such as surface and subsurface flow; erosion and sedimentation of overland as well as channel flows; crop growth for user specified agricultural management practices, and nutrient cycling for various species of nitrogen and phosphorus, among others. The model commonly operates on daily time scale. Spatially, the model subdivides a watershed in to subwatersheds, or subbasins, based on topographic information of the watershed. The subwatersheds could be further classified into spatial modeling units known as hydrologic response units (HRUs) depending on heterogeneity of the land uses and soil types within the subbasins. At the scale of an HRU, watershed variables such as soil types and properties, land use and related management features, weather, and topographic parameters would be considered homogeneous. As a distributed model, a major concern that may arise regarding the practicality of SWAT may be its data requirements. For the U.S., fortunately the minimum data required (e.g., soil, land use, topography, and weather) are commonly

available from government agencies (Nietsch, et al., 2001). For watersheds that lack weather data, the model has a stochastic weather simulator that generates synthetic data based on monthly weather statistics derived from long-term records available from a station geographically located near the watershed. In addition, the model operates on an ArcGIS© platform, which greatly assists in the generation of model input parameters. All these comprehensive features make SWAT an ideal choice for use in integrative watershed management systems.

The Study Watershed and Data

The Morro Bay watershed located on the central coast of California has been used to demonstrate the SA and the calibration methods considered in this study. Two creeks, Los Osos and Chorro, drain this 196 km² watershed into the bay that supports a variety of marine habitats, commercial and sport fishing, shellfish harvesting and recreational activities. The watershed consists of two urban areas, cropland, rangeland, and a wide variety of natural habitats including marsh, oak woodland, riparian, and dunes. The Morro Bay estuary has been impacted by NPS pollution including sediment, bacteria, metals, and nutrients (CCRWQCB, 2002). Sediment has been identified as the major pollutant of concern. To properly simulate hydrologic and water quality fluxes of a watershed, SWAT requires topographic, soil, land use and climate data in addition to observed streamflow and water quality data that is used for calibration. Accordingly, the data that have been obtained for the study watershed include daily rainfall for Morro Bay Fire station from the National Climatic Data Center and other climate data including daily minimum and maximum temperature, wind speed, humidity, and solar radiation obtained from California Irrigation Management Information System (CIMIS) for a station at California Polytechnic State University (Cal Poly) campus. A 10-m resolution DEM and 30-m resolution land use map were obtained from the United States Geological Survey (USGS), and 30-m resolution soil map was obtained from the Natural Resources Conservation Service (NRCS). In addition, streamflow data for three sites in the watershed were obtained from the San Luis Obispo County Public Works, and sediment concentration and streamflow data at additional two sites in the watershed were obtained from previous study conducted by the Central Coast Regional Water Quality Control Board and Cal Poly (CCRWQCB, 2002).

Sensitivity Analysis Methods and Application Results

For distributed watershed models like SWAT that are designed to account for spatial variability of watershed characteristics on hydrologic and water quality outputs, the number of conceptual parameters that need to be calibrated is substantially large, especially when compared to lumped conceptual models where model parameters are averaged over the watershed. The increase in the number of parameters makes calibration of distributed watershed models more challenging. Depending on the characteristics of the study watershed, however, streamflow and water quality outputs may not be equally sensitive to all parameters. SA, which refers to the process of determining the relative importance of individual model parameters on output uncertainty, is an essential component of model evaluation as it helps reduce the number of parameters that need to be calibrated. In this study, a modified form of one-factor- at-a-time (Morris, 1991) that has been implemented in SWAT (Griensven, et al, 2006) and a global SA method that uses sampling based multiple regression approach (Muleta and Nicklow, 2005) are compared in terms of their

effectiveness in screening important SWAT parameters that need to be calibrated for the study watershed.

The one-factor-at-a-time (OAT) method is a local SA method that uses the sensitivity index (S), described mathematically in Eq.(1), as a measure of parameter sensitivity.

$$S_i = \frac{\frac{O(p_1, p_2, \dots, p_i + \Delta p_i, p_{i+1}, \dots, p_N) - O(p_1, p_2, \dots, p_i, p_{i+1}, \dots, p_N)}{O(p_1, p_2, \dots, p_i, p_{i+1}, \dots, p_N)}}{\frac{\Delta p_i}{p_i}} \quad (1)$$

where O is the model output such as streamflow or sediment yield, P_i is model parameter, N is the total number of model parameters, ΔP is the perturbation of the individual model parameter, and i refers to the parameter for which the sensitivity index is being calculated. Morris (1991) introduced an OAT approach that analyzes sensitivity of an individual parameter over its entire range by repetitive application of local SA method. According to the design, multiple parameter sets would be generated by randomly sampling an individual parameter, but keeping all other model parameters at their nominal values. Model outputs will be determined for each of the generated parameter sets and also by changing the parameter value by ΔP . Then the sensitivity index defined in Eq. (1) will be determined thus linking sensitivity of model output due solely to the individual parameter ignoring parameter interaction effects. The design is a significant improvement over the local methods as it assesses output sensitivity over the entire range of individual parameter, not just around a nominal (local) value. The OAT implement in SWAT, known as LH-OAT, is similar with the design described by Morris (1991) except that it uses Latin Hypercube sampling instead of the random sampling used by Morris (1991) thus improving computational efficiency of the Morris method.

For a model with N total parameters, LH-OAT (Griensven et al, 2006) divides each parameter in to user defined K intervals, and generates K Latin Hypercube sample points. Each one of the K Latin Hypercube samples would be altered N times, changing only one of the N parameters at a time. For a Latin Hypercube sample point m , a partial sensitivity index $S_{i,m}$ can be calculated for each parameter P_i , in percentage, as

$$S_{i,m} = \left| \frac{100 * \left(\frac{O(p_1, p_2, \dots, p_i(1 + f_i), p_{i+1}, \dots, p_N) - O(p_1, p_2, \dots, p_i, p_{i+1}, \dots, p_N)}{[O(p_1, p_2, \dots, p_i(1 + f_i), p_{i+1}, \dots, p_N) + O(p_1, p_2, \dots, p_i, p_{i+1}, \dots, p_N)]/2} \right)}{f_i} \right| \quad (2)$$

where O refers to the model functions such as sum of square of errors between the simulated and the observed output, f_i is the fraction by parameter P_i is changed and j refers to the Latin Hypercube point. Each sample point would require $N+1$ model runs. For a model with N parameters where each parameter is divided in to K Latin Hypercube intervals, a total of $K*(N+1)$ model simulations would be needed. Final sensitivity index is calculated for each

parameter by averaging the partial effect described in Eq. 2 over the K runs performed by changing only P_i while keeping all other parameters constant. The final effect is ranked in such a way that the most important parameter is ranked 1 and the least important parameter will take rank N . The attractiveness of the LH-OAT method used in this study is that it integrates the capability of Latin Hypercube sampling to efficiently sample the full range of model parameters with OAT design that ensures that change in model output is precisely attributed to the parameter that was changed.

The global SA method used here (Muleta and Nicklow, 2005) belongs to the sampling based SA methods and it applies a stepwise regression analysis on rank-transformed input-output data pairs to determine quantitative measures of sensitivity. Latin hypercube sampling is used to generate input data from the assigned distributions and ranges, and model outputs are generated by running SWAT for each input data set. All parameters were assumed to follow uniform distribution, and input ranges were assigned based on literature. Relative importance of parameters was determined by the order in which the parameter is selected and entered the multiple regression model, the R^2 value contributed by individual parameters, and the absolute value of the standardized regression coefficient (SRC). Stopping criteria used to limit the number of parameters to be included in the regression model (i.e., sensitive parameters) include the p -value that tests the significance of the additional parameter to the overall performance of the regression model and the improvement achieved in R^2 value between successive regression models. A t -test was also used to test the hypothesis that SRC of the parameter last added to the regression model is different from zero. This global SA method will be referred to as LH-SMR (i.e., Latin Hypercube-Stepwise Multiple Regression) in the remainder of this paper.

To compare the relative performance of these two SA methods, input data collected for the Morro Bay watershed were used. A total of 21 input parameters were considered for streamflow, and each parameter was divided into 20 intervals thus requiring a total of 440 (i.e., $21 * (20+1)$) SWAT simulation for the LH-OAT method. The number of Latin Hypercube samples used for the LH-SMR was kept at 440. For each SA method, SWAT was simulated to determine the sum of square of residuals between the simulated and observed streamflow at Canet Road for each of the 440 Latin Hypercube samples, and the input-output pair was analyzed to determine the relative importance of the 21 input parameters. For the LH-SMR method, the input-output pairs were rank-transformed and stepwise regression was performed on the transformed data.

Results of the two SA methods for streamflow are given in Table 1. For the LH-SMR method, the Table provides the input factors selected at the final step of the LH-SMR model, along with the R^2 of the regression model constructed using the input factor(s) selected at the final step, and the SRC and p -value of each input factor. The threshold values used as a stopping/parameter removal criteria for the difference in R^2 of successive regression models, the p -value based on the test that considers all input factors included to that point (i.e., $b = 0$), and the p -value used to test significance of the individual input factors (i.e., $b_i = 0$) were 0.1 percent, 2 percent, and 5 percent, respectively. For these criteria, the LH-SMR identified nine parameters that play significant role in explaining the uncertainty of streamflow. The relative importance of each of these input factors could be judged using the order in which the parameters were selected, improvement in R^2 that was achieved due to inclusion of the input factor into the regression model and the SRC coefficient of the input factor. For LH-OAT, Table 1 shows the rank each parameter is assigned, the mean and the variance of the sensitivity index calculated using Eq. 2. The rank is assigned based on the magnitude of the mean sensitivity index. For

definition of the input parameters listed in Table 1 and their role in streamflow simulation, the reader is referred to Neitsch et al. (2005) and Muleta and Nicklow (2005).

Comparison of the SA results given in Table 1 shows that overall the two SA methods produced consistent results. Eight of the nine input parameters selected by LH-SMR were also ranked in the top nine by the LH-OAT. However, there is noticeable difference in the relative importance of some of the parameters. For example, Alpha_Bf is ranked 1 by LH-OAT and ranked 9 by LH-SMR. This may be due to high variability of its sensitivity index as determined by LH-OAT (see Table 1). One of the drawbacks of the LH-OAT approach is that, the method generates only relative importance of the parameters. It lacks a quantitative measure that can be used to decide how many parameters to consider for the calibration step making the decision more subjective. LH-SMR overcomes this limitation as it uses several stopping criteria to help decide the cutoff point. Over all, this application results show that either of the two SA methods could be used to identify the most influential parameters that need to be calibrated to improve simulation accuracy of SWAT model. Based on these SA results, the top nine parameters identified by LH-OAT and LH-SMR (i.e., a total ten parameters) have been considered to calibrate streamflow for the study watershed.

Table 1: Sensitivity Analysis Results for LH-SRM and LH-OAT Methods

LH-SMR Results				LH-OAT Results			
Parameter	SRC	P-Value	R ²	Parameter	Rank	Mean	Variance
Cn2	0.7236	0.0000	0.7073	Alpha_Bf	1	1.26563	26.10857
Esco	0.2670	0.0000		Cn2	2	1.25889	1.84401
Sol_z	-0.2192	0.0000		Sol_z	3	0.95305	4.69416
Sol_Awc	-0.1994	0.0000		Esco	4	0.48438	0.49405
Sol_k	0.0940	0.0004		Sol_Awc	5	0.36604	0.02375
Gwqmn	-0.1005	0.0002		Sol_k	6	0.11986	0.00260
Slope	0.0656	0.0133		Slope	7	0.11195	0.00180
Blai	-0.0619	0.0192		Revapmn	8	0.07873	0.12218
Alpha_Bf	0.0517	0.0494		Blai	9	0.06924	0.00586
				Ch_k2	10	0.05013	0.03874
				Epco	11	0.04558	0.00493
				Canmx	12	0.04502	0.00026
				Gwqmn	13	0.02827	0.01568
				GW_Revap	14	0.01490	0.00105
				GW_Delay	15	0.00783	0.00019
				Surlag	16	0.00453	0.00009
				Biomix	17	0.00299	0.00001
				Sol_alb	18	0.00177	0.00001
				Ch_n2	19	0.00105	0.00000
				Tlaps	21	0.00000	0.00000
				Slsbbsn	21	0.00000	0.00000

Calibration Methods and Application Results

Calibration refers to the process of identifying the “best” set of model parameters that would closely match the model simulated outputs and the observed data. Calibration is commonly performed using trial-and-error process where the modeler changes input parameters one-at-a-time and then compares model output with observed data. This manual calibration procedure is time consuming and is less likely to identify optimal set of parameters. To overcome the limitation of this manual procedure, automatic calibration in which optimization algorithms are integrated with simulation models and used to thoroughly search for optimal parameters, is increasingly being used to calibrate comprehensive watershed simulation models such as SWAT (Duan et al. (1992), Muleta and Nicklow (2005), Tolson and Shoemaker (2007)). In this study, four optimization methods: Genetic Algorithms (Holland, 1975), Shuffled Complex Evolution-University of Arizona (Duan et al., 1992) as currently implemented in SWAT, Parameter ESTimation (PEST) (Doherty, 2004), and Dynamically Dimension Search (DDS) (Tolson and Shoemaker, 2007) have been used to automatically calibrate streamflow for the Morro Bay watershed. The methods are compared in terms of their effectiveness in improving accuracy of the simulated output and in terms of their computational efficiency. The four optimization methods have been selected because of their popularity in calibrating watershed simulation models and based on results of comparative studies that have been reported in the literature.

Genetic Algorithms (GAs) are heuristic search algorithms that apply the principle of genetics and the Darwinian theory of natural selection and survival of the fittest to optimization. As a subset of evolutionary computation, GAs require no derivative information about the objective function or constraints and has been practically proven to work well on nonlinear, nonconvex, and multimodal problems. Though not ultimately guaranteed to locate global optima, GAs search a wide portion of the solution space and, thus, have a better capability of locating optimal solutions. In fact, the majority of GAs literature consistently demonstrates an ability to identify global or very near global optima for a range of complicated problems. GAs have been successfully applied to optimization of hydrologic models (Wang (1991), Muleta and Nicklow (2005)). In this study, continuous GAs have been developed and integrated with SWAT for calibration of streamflow.

Shuffled Complex Evolution-University of Arizona (SCE-UA) (Duan et al., 1992) is probably the dominant automatic calibration algorithm for hydrologic models during the past decade. The SCE-UA currently implemented in SWAT starts by randomly generating initial population of potential solutions from the feasible parameter space. The initial population is then partitioned into complexes, each containing $2N+1$ points, where N is the number of parameters to calibrate. The Simplex method (Nelder & Mead, 1965) is used to evolve and guide each complex independently to locate potential optimal solutions. The population from all complexes is periodically shuffled and new complexes are formed so that the information gained by the previous complexes is shared. The evolution and the shuffling steps continue until prescribed convergence criteria are reached. The default stopping criteria implemented in SWAT is used in this study.

Dynamically Dimension Search (DDS) (Tolson and Shoemaker, 2007) has been developed to improve computationally efficiency of calibrating comprehensive watershed simulation model. DDS is a simple stochastic search method that starts by globally searching the feasible region and incrementally localizes the search space as the number of simulation approaches the maximum allowable number of simulation. Progress from global to local search

is achieved by probabilistically reducing the number of model parameters modified from their best value obtained thus far. New potential solutions are created by perturbing the current solution values in the randomly selected model parameters only. The perturbations magnitudes are randomly sampled from a normal distribution with a mean of zero. Tolson and Shoemaker (2007) compared DDS and SCE-UA to calibrate SWAT and found that DDS required only 10-15% of model evaluations for the same level of performance. In this study, source code of the DDS algorithm described in Tolson and Shoemaker (2007) has been obtained from the first author and has been integrated with SWAT to calibrate streamflow for the study watershed.

PEST (Doherty (2004), Gallagher and Doherty (2007)) is a gradient-based optimization method that uses the variant of Gauss-Marquardt-Levenberg to identify optimal solutions. When compared to global optimization methods such as GAs and SCE-UA, PEST generally requires fewer model runs to solve a minimization problem, but is more susceptible to be trapped at a local minimum. This problem may be minimized by starting several Gauss-Marquardt-Levenberg calibration runs from different points in parameter space which are selected in a manner that minimizes the chance of finding the same local minimum twice (Skahill and Doherty, 2006). PEST is becoming increasingly popular for calibration of hydrologic model partly because it is a model-independent parameter estimator as it communicates with the model being calibrated through the model's own input and output files without requiring any coding or changes to the model. Such standardized input/output approaches to model evaluation tools are regarded as the future of models and model evaluation tools (Mattott et al., 2009). In this study, inputs and outputs required by PEST were prepared and used to calibrate SWAT for the study watershed.

The four optimization methods were used to calibrate streamflow for the Morro bay watershed. For GAs, SCE-UA, and the DDS algorithms, a maximum of 5000 model evaluations was used. PEST does not use maximum iteration as stopping criterion. Five calibration runs were made for each of the four methods which is a total of twenty calibration runs. Cognizant of the fact that performances of the calibration methods largely depend on values used for the user specified parameters of the respective algorithm, all such parameters were set to their default or commonly used values. Sum of square of residuals between model simulated and observed outputs was used as objective function. Daily streamflow from 1995-1998 collected at two sites in the watershed were used to calibrate the model, and a warm up period of six months (07/1994-12/1994) was used to diffuse the effect of initial conditions. Calibration results obtained by the four methods are summarized in Table 2. Graphical comparison of the results obtained using PEST for Canet Road, one of the two stations, is given in Figure 1. Table 1 shows that PEST and DDS outperformed SCE-UA and GAs in terms of effectiveness as well as efficiency. PEST needed only 120 evaluations to produce the results shown in the table. DDS was the second efficient as it needed less than 2000 simulation to converge. Based on these results, one can conclude that PEST is by far the most efficient and it also performed better than GAs, SCE-UA and DDS. All four methods produced good results compared to past attempts at modeling watersheds in arid and semi-arid regions using SWAT (Van Liew et al., 2007). More detailed comparative study is being done to confirm the results reported here.

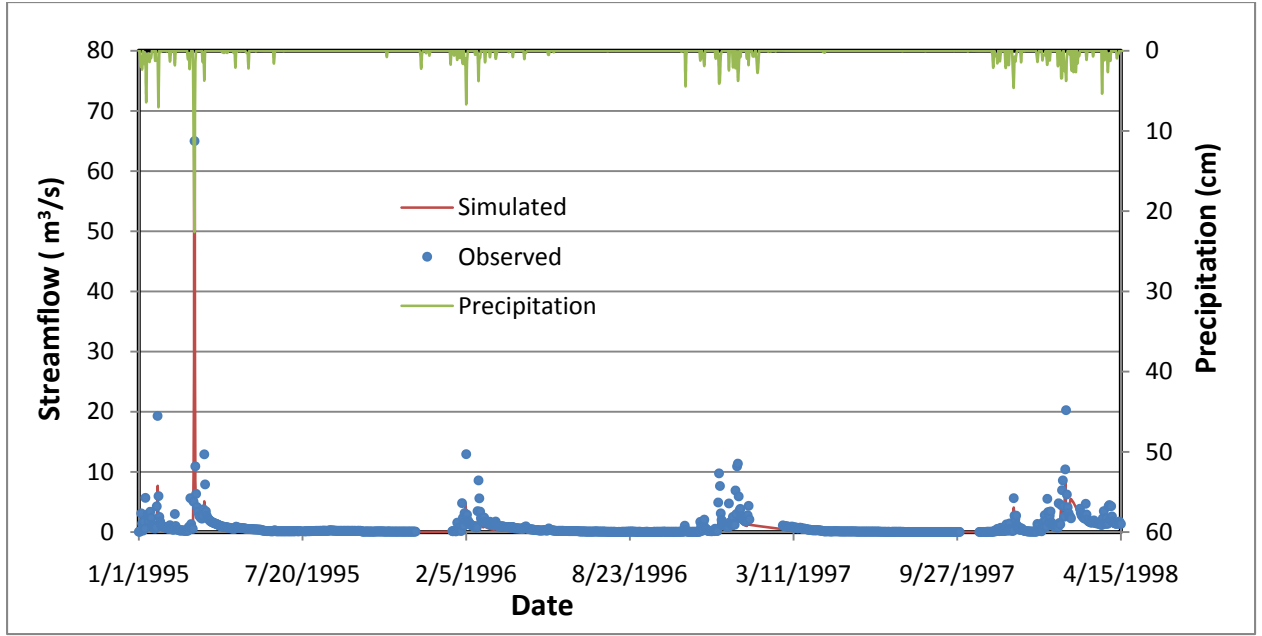


Figure 1: Comparison of Simulated and Observed Streamflow for Canet Road

Table 2: Calibration Results Obtained Using Four Optimization Methods

Statistics	GAs	SCE	DDS	PEST
<i>Mean Sum of Square of Residuals</i>	3542.98	4022.1	3420.62	3419.0
<i>Average E_f</i>	0.5272	0.5657	0.5419	0.5466

Conclusions

This study presents model evaluation steps that have been undertaken to simulate streamflow for Morro Bay watershed using SWAT. Two sensitivity analysis methods, LH-OAT and LH-SMR, have been used to determine the most influential parameters of SWAT that need to be calibrated. LH-OAT and LH-SMR identified similar parameters as influential parameters with the same number of model evaluations indicating that both methods can be used for such applications. Four automatic calibration methods have also been compared in terms of their efficiency and effectiveness. The calibration comparison shows that PEST, a gradient based local search method outperformed global optimization methods including GAs, SCE-UA, and DDS. More detailed comparative study will be needed to confirm the results reported here.

References

- Arnold, J.G., Williams, J.R., Srinivasan, R., and King, K.W. (1999). SWAT: Soil and Water Assessment Tool. U.S. Dept. of Agriculture, Agricultural Research Service, Temple, TX.
- Beven, K.J. and Freer, J. (2001). "Equifinality, Data Assimilation, and Uncertainty Estimation in Mechanistic Modeling of Complex Environmental Systems Using the GLUE Methodology." *J. of Hydrology*, 249: 11-29.
- CCRWQCB (2002). "Morro Bay Total Maximum Daily Load for Sediment (including Chorro Creek, Los Osos Creek and the Morro Bay Estuary)." State of California Central Coast Regional Water Quality Control Board, San Luis Obispo, California.
- Doherty, J. (2004). PEST: Model-Independent Parameter Estimation, User Manual, 5th ed., Watermark Numer. Comput., Brisbane, Queensland, Australia.
- Duan, Q., Sorooshian, S., and Gupta, V.K. (1992). "Effective and efficient global optimization for conceptual rainfall-runoff models." *Water Resour. Res.*, 28(4): 1015-1031.
- Gallagher, M., and J. Doherty (2007). Parameter estimation and uncertainty analysis for a watershed model, *Environ. Modell. Software*, 22(7), 1000–1020.
- Griensven, A. v, T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *J. Hydrology*, 324, 10–23.
- Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.
- Matott, L. S., J. E. Babendreier, and S. T. Purucker (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools, *Water Resour. Res.*, 45, W06421, doi:10.1029/2008WR007301.
- Morris, M.D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics* 33(2).
- Muleta, M.K. and Nicklow, J.W. (2005). "Sensitivity and Uncertainty Analysis Coupled with Automatic Calibration for a Distributed Watershed Model." *J. Hydrology*, 306: 127-145.
- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., and Williams, J.R. (2005). *Soil and Water Assessment Tool Theoretical Documentation*. Version 2005. Grassland, soil and water research service, Temple, TX.
- Skahill, B.E. and J. Doherty (2006). Efficient accommodation of local minima in watershed model calibration, *J. of Hydrology*, 329, 122– 139, doi:10.1016/j.jhydrol.2006.02.005.
- Tolson, B. A., and C. A. Shoemaker (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, 43, W01413, doi:10.1029/2005WR004723.
- U.S. EPA (2002). Guidance for quality assurance project plans for modeling. EPA QA/G-5M. Report EPA/240/R-02/007. Washington, D.C.: U.S. EPA, Office of Environmental Information.
- Van Liew, M.W., Veith, T.L., David D. Bosch, D.B. and Arnold, J.G. (2007). Suitability of SWAT for the Conservation Effects Assessment Project: Comparison on USDA Agricultural Research Service Watersheds. *J. Hydrologic Engineering*, 12:2(173).
- Wang, Q. J. (1991). The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models. *Water Resour. Res.*, 27(9), 2467–2471.